

## 第10回ゲノム医療協議会 議事概要

■日時：令和4年12月23日（金）13時00分～15時03分

■場所：中央合同庁舎第4号館12階全省庁共用1208特別会議室

■出席者：

- 議長：松尾泰樹 内閣府科学技術・イノベーション推進事務局長  
議長代行：西辻 浩 内閣府健康・医療戦略推進事務局長  
構成員：木村直人 文部科学省大臣官房審議官（研究振興局及び高等教育政策連携担当）〔代理〕（Web参加）  
浅沼一成 厚生労働省大臣官房危機管理・医務技術統括審議官（Web参加）  
中谷祐貴子 厚生労働省健康局がん・疾病対策課長〔代理〕（Web参加）  
城 克文 厚生労働省大臣官房医薬産業振興・医療情報審議官（Web参加）  
下田裕和 経済産業省商務・サービスグループ生物化学産業課長〔代理〕（Web参加）  
上野裕明 日本製薬工業協会 副会長  
春日雅人 国立研究開発法人日本医療研究開発機構（Web参加）  
ゲノム・データ基盤プロジェクトプログラムディレクター  
菅野純夫 国立大学法人千葉大学未来医療教育研究機構特任教授  
日本学術会議基礎生物学委員会・統合生物学委員会・基礎医学委員会合同ゲノム科学分科会委員長（Web参加）  
高木利久 富山国際大学 学長（Web参加）  
門田守人 一般社団法人日本医学会連合会長（Web参加）  
山口 建 静岡県立静岡がんセンター総長（Web参加）  
米村滋人 東京大学大学院法学政治学研究科 教授（Web参加）  
参考人：岡田随象 大阪大学大学院医学系研究科遺伝統計学/東京大学大学院医学系研究科遺伝情報学 教授  
鎌谷洋一郎 東京大学大学院新領域創成科学研究科  
メディカル情報生命専攻複雑形質ゲノム解析分野 教授  
小林啓之 株式会社 Preferred Networks 執行役員  
ヘルスケア&ウェルネス担当 VP  
田端佑介 株式会社NTTデータ 技術革新統括本部技術開発本部  
イノベーションセンタ シニアエキスパート  
橋本和洋 株式会社Mitate Zepto Technica 取締役副社長  
松岡 豊 国立研究開発法人日本医療研究開発機構  
データ利活用基盤整備推進準備室利活用推進課長（Web参加）

オブザーバー： 笠貫 宏 内閣府健康・医療戦略参与  
中釜 斉 内閣府健康・医療戦略参与 (Web 参加)  
森下竜一 内閣府健康・医療戦略参与 (Web 参加)  
三島良直 国立研究開発法人日本医療研究開発機構 理事長  
(Web 参加)

議 事：

- 1) ゲノムデータのユースケースについて
- 2) 全ゲノム解析等実行計画の進捗について
- 3) 健康・医療研究開発データ統合利活用プラットフォーム事業の進捗について
- 4) その他

■概 要：

○長野次長 それでは、時間が来ております。米村先生がもうしばらくしたら入られるのではないかと思いますので、始めさせていただきます。ただいまより第10回「ゲノム医療協議会」を開会いたします。

本日は御多忙の中、御参集いただきまして、どうもありがとうございます。

私、本日の議事進行を務めさせていただきます健康・医療戦略推進事務局の次長の長野でございます。どうぞよろしく願いいたします。

今回もオンラインとの併用で開催となっております。

本日、構成員は全員御出席の予定となっております。米村構成員が後ほど入られる状況です。

議事1) ではゲノムデータのユースケースに関してとなっておりますが、本日はアカデミアより岡田先生、鎌谷先生、また、企業からPreferred Networksの小林様、NTTデータの田端様、Mitate Zepto Technicaの橋本様より御発表いただく予定でございます。

また、オブザーバーとしては、笠貫参与、中釜参与、森下参与、三島AMED理事長にも御出席いただいております。

最初に、議長であります松尾科学技術・イノベーション推進事務局長より御挨拶をお願いいたします。なお、松尾議長は公務のため御挨拶の後退席いたしますので、よろしくお願いいたします。

○松尾事務局長 皆様、どうもこんにちは。ただいま御紹介いただきました科学技術・イノベーション推進事務局の松尾でございます。

本日は御多忙の中、御参集いただきまして、本当に心から感謝申し上げたいと思います。

今、長野次長からもございましたが、今回の協議会におきましても前回に引き続きましてゲノムデータのユースケースについて我々は勉強させていただきたいと思っております。特に今回5名の先生方、アカデミアの先生お二人、企業でやっておられる先生方3名ということで、幅広く意見交換できればと思っております。そこで得られた新たな知をどの

ように還元していくかについて、ぜひとも一緒に考えていきたいと思っております。

特に今年度から全ゲノム解析等の実行計画におきまして医療の実装に向けた準備が進んでいると伺っておりますが、その詳細を御報告いただけるということでございます。今回はこの実行計画におきまして研究成果をいかに利用するかということについて、ゲノムの医療推進に不可欠なELSIの取組についてもしっかりと御報告いただきたいと思っております。これらのユースケースを受けまして、データの利活用の実際のニーズを踏まえたCANNDsの構築・運用に生かせるよう実のある議論を期待しておりますので、何とぞよろしくをお願いいたします。

構成員の皆様方におかれましては、個別化医療、個別化予防の実現に向けたゲノム医療のあるべき姿をしっかりと想定した上で、このゲノムデータの利活用をどのように進めていくべきかという観点で御示唆いただければありがたいと思っております。

長野次長からもありましたけれども、早めに抜けてしまいますが、各省庁におかれましては、本日の議論を踏まえてしっかりと取組を検討するようお願いをしたいと思います。何とぞどうも今日はよろしくをお願いいたします。

○長野次長 ありがとうございます。

それでは、議事に入る前に、本日の資料について事務局より確認をさせていただきます。

○笠松参事官 本日の資料は、議事次第に記載しておりますとおり、資料1-1から1-5、資料2、資料3-1、3-2並びに参考資料1でございます。事前にメールでお送りしておりますので、御確認くださいようお願い申し上げます。

○長野次長 本会議は公開とし、希望する記者、傍聴者の方に入室いただいておりますのでございます。

なお、本日配付の資料及び逐語ベースの議事概要を後日公開させていただきますので、よろしく申し上げます。

冒頭のみカメラ撮影を認めますが、よろしいでしょうか。

それでは、議事1)に早速入らせていただきます。まず、5名の方からそれぞれの専門の見地よりゲノムデータのユースケースについて御発表いただきます。それぞれ10分程度ということで御発表いただいて、最後に総合討論という形で意見交換の時間を取らせていただきたいと思っております。

最初に、岡田先生からよろしくお願いいたします。

○岡田参考人 大阪大学の岡田でございます。このたびはこのような機会を賜り感謝申し上げます。

それでは、早速私の発表を始めさせていただきます。資料1-1「遺伝統計学の国際展開戦略」という名でお話をさせていただきます。

第1枚目は私の簡単な自己紹介となります。現在東京大学に戻ってまいりまして、遺伝統計学を中心に疾患病態・ゲノム創薬・個別化医療に取り組んでいるということでござい

ます。様々なオミックス解析を実施しておりますが、興味のある先生方におかれましては、後日私どものホームページ等を御覧いただければ幸いです。

続きまして、ゲノム配列情報取得コストの劇的な低下についてであります。これは皆さん御存じのように、現在ゲノム研究というものは非常にコストが低くなり、参入障壁が下がっていると。どの国でも誰でもできる時代になった。これが大事であります。SNPマイクロアレイは40ドルぐらいですね。UKバイオバンクが50万人規模で現在全ゲノムシーケンスもつくり、かつそれを世界中に分譲しているということで、必要とされる規模感、また、データアクセスでゲームのやり方が変わってしまったということが大きいところあります。このように大きくなりますとたくさんの方が見つかっておりまして、大きいものを使ってたくさんの方の感受性遺伝子を見つけるということはルーチンのことになってしまいました。

続きまして、国際コンソーシアムであります。このような形になってくると何が大事か。これはいかに世界に散らばっているリソースをまとめてコンソーシアムとして成果を出し、次につなげていくかという国際イニシアチブの獲得競争に入っているわけです。ここに示しましたのは、いわゆる誰でも知っているみんなが使う有名なコンソーシアムであります。こういったコンソーシアムがつくったデータがスタンダードとなって、これがガイドラインをつくる基になっていくわけですね。このようなプロジェクトがたくさん進んでおります。ほとんどというかほぼ全てアメリカかイギリスが主導しているものであります。ですから、データの規模だけでなく国際的ガイドラインの獲得競争においても彼らは非常にうまくやっているなということなのです。これはぜひ我が国も食い込んでいく必要がございます。

次は国際バイオバンク連携ということで、といたしまして、日本にも非常に素晴らしいリソースがございます。これは本邦が誇るバイオバンク・ジャパン、UKバイオバンク、FinnGen等の国際バイオバンク連携で、50万人を超えるリソースで200以上の形質を解析するとたくさんの方が見つかりましたよということでありまして。こういったことが現在世界中でルーチンで行われているわけですね。私どもとしましては、この解析結果は全て一般公開して誰でも触れることにより、それによって私たちの価値を上げていくといった戦略を取らせていただいております。

続きまして、コロナの話であります。こういうコンソーシアムは我が国でも頑張っているよという例でございます。今回のパンデミック、2年半ほど経過いたしておりますが、私どもは完全なアカデミア発の有志連合としてタスクフォースを立ち上げました。結果といたしまして、100以上の病院から6,000名以上の感染者のゲノム・RNA・血漿・臨床情報を集め、これはアジア最大の生体レポジトリとなっております。これを行いまして、例えば重症化遺伝子の同定ですとか、重症化のメカニズムの解明ですとか、国内にあっても皆が力を合わせることで新しいエビデンスを出していくことができることを示していきたいかと考えております。重症者に関する予測もかなり精度が高くなってきておりまして、コロ

ナが次に起きるか分かりませんが、10倍以上の非常に高い精度での重症化の事前の絞り込みが可能となっております。

続きまして、All Handai、これはドメスティックな視点であります。大阪大学におきましても私たちは頑張っています、これは若手の研究者が大学内で部局を超えてみんなが持てるものを合わせようということで、みんなで頑張っております。シングルセル解析を中心にみんなのリソースを集約して、これも結果が出つつあります。つまり、国際レベルであれ、国内レベルであれ、学内であれ、一人がやるというよりはかにかにいいチームワークで成果を創出していくかということが現在鍵となっているわけでございます。

このような紹介の後にはゲノム個別化医療の話をしていただきます。皆様御存じだと思います。Precision Medicine Initiativeというものがあまして、ゲノム個別化医療ですね。ゲノムのついた個別化医療の時代が幕を開けているわけでもあります。そして、有名なPolygenic Risk Score、ゲノム情報に基づいて個人のリスクを事前に定量化しようということが進んできております。その予測できるかどうかということに関しては、これはできるようになっています。例えば非常にリスクの高い心筋梗塞においても、上位0.5%の集団が5倍ということで、いわゆる健康診断ベースで出てくるリスク因子よりも高い確度とエフェクトを持った予測が可能となっております。

続きまして、このポンチ絵であります。将来どうなっていくか。既にこれはなりつつあるわけではありますが、ゲノムに基づく予測、これは人生の早期にできることにメリットがあるわけです。したがって、加齢に伴い集団中で高頻度に発症していく生活習慣病のような病気においては、人生早期に基づきジャームラインによるPRSの層別化を行った上で、これに対して様々なオミックス情報を合わせながら前向きに予防していきましょうという話になっております。ここで大事なことは、予測にかかる費用は実は非常に安いということです。ゲノムモデルの構築は1人1万円かからないわけです。うまく回すことができれば、これは医療経済的にも非常に効率化できることは間違いないかと考えているところであります。

次は少し別の紹介になります。Polygenic Risk Scoreを予測するだけではなくてほかの使い方もできますよという方法になります。これはPRSを、細かいところは省きますが、集団中における生まれながらのみんなの病気のリスクを考えることで、どういうバイオマーカーが結局寿命を決めているかという健康長寿のバイオマーカーはなかなか見つからないのですけれども、それを見つけることにゲノムを使ったという方法であります。これは高血圧、肥満が寿命を短縮しているというある意味当たり前の結果が出てきておりますが、因果関係をもってこういうことが分かる、つまり、高血圧、肥満をなくせば寿命が延びることを示せたこと、これがゲノムのパワーであります。このようにPRSを疫学研究に組み込んでいくことで、より因果関係と介入の効果を示した成果を出していくことができます。

続きまして、ガイドラインです。こうなってきますと、PRSの予測というのもできるしやるべきであります。ここもガイドラインを策定して、世の中はものをつくった者よりもル

ールをつくった者が勝つという前提があるわけです。私どもは例えば国際ゲノム政策提言組織でのホワイトペーパーの発表ですとか、そういうものを通じたベストプラクティスの提言というサイエンスというよりは規格の決め方ですね。こういう中に私たちの国も入っていかなくてはいけない時代が来ているわけでありませぬ。

駆け足ですみません。次は創薬の話となります。何で創薬が大事かという話です。ただ、疾患ゲノム情報が創薬プロセスを効率化する、これに関してはエビデンスはほぼ固まっております。幾つか表がありますが、ステージに応じてエビデンスが増すとか、ゲノムの根拠があるものばかり承認されているとか、これは既知の事実となっております。製薬企業からすると、問題はいかにゲノム情報、臨床情報にアクセスし、それを彼らのパイプラインに生かすかという実務面になってきております。産官学民によるバイオバンク運営とありますが、そこでうまくいきましたのがフィンランド、FinnGenでありまして、製薬企業等のパートナーよりお金をもらいつつ、データを渡しつつ、ウィン・ウィンな関係で後発のバイオバンクでありながらかなり国際的な存在感を出すことに成功したのがこのFinnGenのモデルであります。

次のスクリーニングの話は私の研究の少し紹介となります。アカデミアとしてもゲノムデータ解析をやったときにどのように薬の種を見つけていくべきかということは、方法論として開発しリードしていく必要があります。私たちの研究室ではGREPですとか、Trans-Pharとか、そういったソフトウェアの開発を進めております。私も2年間だけ医者をやっていたことがありまして、その経験がこういうときには少し生きるかなどと思いつつながら研究をしております。これもどうやっていくかというベストプラクティスの提唱と方向性の策定が大事なのです。このようなことに関しましても国際コンソーシアムの枠組みで私どもはさせていただいております。

次のこれからのゲノム創薬、これはこういった枠組みをコンピューター上だけで終わらせなくて、vivoからat bedでどうつなげていくかが大事であると。iPS細胞を使った分化ですとか、モデル生物とか、本邦に強い技術がありますので、現在こういった専門の先生方とコラボをしながら、私どものコンピューター上の話がどう世の中に実装していくかという研究を進めております。

次は人材育成です。結局人が増えなければ成果は増えないということで、本邦はこの遺伝統計学に関する人材不足が際立っているわけですが、私はバイオインフォマティクスの「夏の学校」というものを行っております。もう7回行いました。今年度は300名の方に参加していただきまして、高校生から名誉教授の先生方に参加いただくという会になっております。これは本当に簡単なことしかしておりませぬ。メスしか握ったことのない先生でもできるようになる、これが目標でありまして、実際にみんなできるようになります。これを使ってぜひインフォマティクスのパワーの底上げをしていきたいと思っております。大阪大学医学部ではこれは実習として採用がされていまして、今年から東京大学医学部においてもこういったことをスタートさせる予定でございます。

次から3枚のスライドが、私が考えるメッセージであります。現在の実情は何か。過去20年間はアジア人にデータがないという貴重性の中で我が国は非常によい成果を出してきました。しかし、どのアジアの国でもバイオバンクをつくれる現在、もうかじ取りは変えなければいけないのですね。何よりデータの量だけではなく国際的なコミュニティ、ガイドライン策定における存在感の低下が著しいというわけであります。

さらに、解析技術の格差が開いております。シーケンスに関してはサンプル数のレベルでゼロが2つぐらい乖離しています。他のオミックス技術がどうかというと、日本は昔はこれは強かったのですが、ベンチャー、スタートアップというところの遅れがありまして、最新鋭のオミックス解析技術が日本でできるようになるまでのタイムラグがどんどん延びていて、ここが私たちがデベロッパーからコンシューマーになってしまっているという如実な点なのです。

そして、遺伝統計学の若手人材の不足解消が急務であります。お医者さん出身のデータ解析研究者は結構増えております。そして、お医者さんによるオミックス解析の融合研究は結構成果が出ています。これは日本のMDが非常によく働くという延長線で非常によく研究をしているからです。すばらしいのですが、これには限界があります。私たちのようなお医者さんがちょっと統計をかじったので通用する時代ではありませんので、数理科学、情報科学、統計学の専門家の参入がぜひとも必要で、彼らが入ってきてくれる余地をつくらなければいけません。そして、ここに大事なのがデータの公開・共有・解析基盤の遅れですね。最先端のデータに触れられなければ彼らは参入障壁として入ってきてくれないわけで、さらなる人材不足が続くというフィードバックが続いております。

次は戦略です。では、どうしたらいいのでしょうか。これは広い話になります。ただ、私は医学部、生命科学科における実験もいいですが、キャリア早期にインフォマティクスが全員できるようになる、これは必須だと思います。現在インフォマティクスを要さないライフサイエンスはなくなりましたので、これはもう大学の基礎教育でやってしまうと。

そして、どんどん海外に送り出すわけですね。国際コミュニティにいるためにはそこに人がいなければできません。海外留学、ポスドクもそうですが、博士課程から優秀な子は海外の大学院に行かせる、そういう時代であり、これは制度面、財政面でのサポートが必要です。

そして、この「アジアの雄」を「真剣に」目指すといいますが、正直に申し上げて現実的な目標は、世界一を目指すというよりはアジアの中で一番をどう目指すかに僕はシフトしていると思います。規模では中国には勝てませんし、リーダーシップではシンガポールになかなか勝っていませんね。これは強く感じます。なぜか。実はこれは英語化です。公用語が英語の国にはどうしても負けてしまうのです。ですから、せめて学会・大学院の英語化を推進して、大学ランキングもあれは何なのという意見もありますが、ああいうところでも英語化していないということがランキングの足を引っ張ってしまうわけですね。こういうところが実は効く時代となってきております。

そして、ウェット・ドライの解析、これはもうラボで行う時代ではありません。コアファシリティー化して少なくとも施設内では一元化、できれば施設間でも一元化するようにして効率化を図っていかなければいけませんし、オミックス解析・知財出願のサポートも大事です。

そして、こういう非常によい研究ができる分野に数理、情報、統計の専門家の人に入っただけ、彼らが気持ちよくキャリアパスを形成していけるようにしていかなければいけません。

そして、大規模ゲノムデータ・バイオバンクデータの促進ですね。データ公開・分譲の遅れが人材の参入障壁になっているというのは申し上げたとおりです。はっきり申し上げて、大きなところが見本を見せる、これが全てであります。私は個人の研究者としては日本でかなりデータをデポジットし続けてきたほうではありますが、もっとみんな出せるのではないか、大きいところは出せるのではないかということでもあります。そして、バイオバンクにおけるデータ所有権の属人化ですね。〇〇先生のものということが起きているわけですが、そういうことをなくして、これはもう国のものであり組織のものである、そのようにしていかなければいけません。データに関しましても中間データだけ公開して終わりという例が非常に多くて、元データからですと。GWASといたらintensity data/idadtファイル、全ゲノムシーケンスはfastqファイル、ここまできちんと公開しなければなかなか公開したといっても研究は進まないわけであります。

最後は今後の展望です。これまではデータの不均衡がある中で、いかに持たざる者から持つ者の側によるかという戦略が正しかったのですが、もうそれは無理です。誰でも世界最高峰のデータにアクセスできる時代になってしまったという認識の下、違った戦い方を僕らはつくる必要があります。これはもう医学だけではありません。情報、数理、生物、疫学、いろいろな人がチームをつくり、そして、どんどん新しい、これは非常に僕らが予測するより常に速く進む解析技術の進歩をどんどんキャッチアップして、個別の技でパフォーマンスを出すことと総合力で勝ちに行く。これを2つしなくてはいけません。

領域に関しましては、解明や予測は結構できるようになってきました。今後は制御に行くのです。予測結果を伝えてフィードバックし、どう制御していくか。これはまた違う学問への変革が迫られています。データ解析に関しましては、思ったよりお金がかかるという事実もありますので、ここは予算配分によるコスト認識も大事かと思えます。海外資本の柔軟活用も大事です。Finngen Partners、あとUKバイオバンクは5万人のプロテオームを出して、今、非常にやっています。「UK Biobank Pharma Proteomics Project」ですね。つまり、彼らくらい潤沢な資産と資金があるところでも海外資本の柔軟な活用を行っているわけですから、私たちもどうしてもそういったことを考える必要があると思えます。

そして、海外研究者への開放も必要であります。この「安全な」とつけましたのは、国内と海外を少し分ける必要がありまして、何でもかんでもあげるといっただけではありません。しかし、開放しなければネグレクトされてパッシングされてしまうという事実もあり



ますから、これはいろいろなGenomics Englandの先行例がありますけれども、いかにイニシアチブを取るのにベストかということを外カムに海外への安全な開放も検討していただけたらいいのではないかと考えております。

すみません。少しオーバーしてしまったかもしれませんが、以上でございます。

○長野次長 ありがとうございます。

岡田先生からは遺伝統計学を主軸としながら、それを超えた形で幅広い御説明をいただきました。ありがとうございます。

続きまして、オミックスデータの創薬への応用ということから、鎌谷先生より御発表いただきます。どうぞよろしく願いいたします。

○鎌谷参考人 よろしく申し上げます。東京大学の鎌谷でございます。

最近、岡田先生と私とでこういったところにお呼びいただくことが多くて、岡田先生の御発表がすごいので私は多少ストレスに感じているのですけれども、足元でこういったところで頑張っていこうかなというところをお話しさせていただきたいと思っております。

2枚目の資料を見ていただきますと、こういったところで創薬をゲノムからやろうと。それはなぜかという、成功したからだということです。具体的には、これは2003年ですね。PCSK9という遺伝子の変異によって高コレステロール血症が起こると分かると、では、このPCSK9の機能を落とすことで治療できるのではないかとということで、薬ができる。そのできた薬が実際のブロックバスターになっているということです。左側の研究はフランスのクリスティーヌ先生が行われた研究ですけれども、実際にクリスティーヌ先生は今、PFMGというフランスの全ゲノム実行計画を引っ張っていらっしゃるしやったりしまして、こちら辺が実際に世界でもゲノムオミックスの大規模研究を引っ張るドライバーになっているということです。右側の研究はSGLT2という遺伝子が機能を喪失すると血糖が下がるということで、そうすると、やはりSGLT2の機能、たんぱく質の機能を落とそうと、受容体の機能を落とそうという戦略になるというシンプルなお話から始まっています。

次に行っていただきますと、こういったことは単一遺伝子疾患の変異によることでは成功するということは十分なコンセンサスがございすけれども、もっとありふれた広いたくさんの人がなるような病気によってうまくいくかということですが、岡田先生と全く同じ図を用意してしまいましたけれども、うまくいくという根拠はどんどん出てきておまして、上側の研究は過去の製薬企業が持っている化合物について調べたところ、ありふれた疾患のゲノム解析が成功した薬をちゃんと予想しているということでしたが、下側の今年発表されたものによると、実際に去年認可に至った薬がゲノム解析結果の遺伝子と疾患の関係性としてしっかり当たっているものがたくさん認可されていることが分かってきたということです。

その次に行っていただきますと、そうはいっても過去に出てきた薬は当たっていたというのは多くあるのですけれども、実際にゲノム解析をきっかけに創薬をしたもので成功したもの、PCSK9などは成功していますけれども、ありふれた疾患ではどうかということでは

うと、いろいろ皆さん知っているところもあるかもしれません。私の知っているところですと加齢黄斑変性症で、この疾患は多因子疾患ですけれども、C3という遺伝子の関連が発見されていますけれども、このC3の遺伝子をターゲットとした薬が現在優先審査指定されている状況と把握しています。この薬は臨床試験においても遺伝的関連解析の結果、この遺伝子をターゲットとした創薬を行ったと明記している薬ですので、こういった方向で海外の製薬企業がゲノム研究に大きくお金を出しているといったところは、ここら辺の経過を見てのことという状況になります。

次のところに行ってくださいますと、そう言いますけれども、海外で非常に大規模なゲノムデータが出ている状況で、日本でゲノム解析をやって何か意味があるのかと。これは実際は2013年頃だったと思うのですけれども、海外で私はこうやって、日本人はゲノムデータが違うから新しい関連遺伝子が出ると思うと言ったときに、おまえは本気でそんなことを考えているのかと言われたことがありましたけれども、それは出ています。東京大学の山内先生、門脇先生が主導されました日本人の2型糖尿病のゲノム解析では、GLP-1Rという遺伝子の関連を見つけました。これは欧米では見つかっていない関連でして、実際にこのGLP-1Rは既に先に分子生物学的研究で分かっていた遺伝子でして、薬になっていますけれども、日本人でだけしか見つからなかったと。結果としてみるとそういった遺伝子が見つかることは分かっています。

次のページに行ってくださいますと、去年「Science」にGPR75という遺伝子が肥満に関係する遺伝子として報告されていました。これまで肥満に関係する遺伝子は遺伝子機能が落ちると太ってしまうというものが報告されていたのですけれども、今回この「Science」の研究ですとGPR75という遺伝子機能が落ちると痩せることが初めて発見されて、そうすると、たんぱく質は機能を上げるよりも下げるほうが簡単ですので、薬によって痩せさせることができるのではないかということで、これも創薬プロセスの開発プロセスに入っていると伺っています。実はこの遺伝子、私たちは先に発見してまして、2017年の時点でバイオバンク・ジャパンのBMIの肥満のゲノム解析をやったときにGPR75を発見できてはいました。ですから、結局ここでの問題は何かということ、遺伝子を発見してそれをどう生かすか、どこにお金をかけるかということに関してあちらは明確な目的意識を持ってやっていて、実際に動いていて、我々は動けなかったといったところに私の反省もありますし、国内でもこういった我々自身が持っているデータが本当に重要なものだということを理解していきたいと思っています。

ここまで、最初に単一遺伝子疾患から薬は成功するというお話をして、次に多因子疾患で1つの遺伝子がちゃんと見つかったときにうまくいきそうだということをお話ししましたが、ただ、実際には多因子疾患においては遺伝子をきちんとターゲットとして遺伝的結果が得られることはまれでして、多くは遺伝子の外に見つかる、ヒトゲノム上の98.5%ぐらいの場所は遺伝子ではないので、そこでたくさん見つかる場所がターゲットを絞りにくいということで言われておりました。

次のところで、そういうわけで実際にほとんどのありふれた疾患でバリエーションを見つけると、ここに何が描いてあるかということ、赤いところが、遺伝子が当たっているものですが、他のグレーのところは全部遺伝子に当たっていないということで、そうすると、ゲノムの結果としてターゲットで創薬をしようにも何の遺伝子を使えばいいかわからないというところでした。

そういったわけでその次のところ、シンプルにセントラルドグマから攻めて、遺伝的に変化が起こったら生物で実際に何が起こるかを解明しなければいけないとここ10年みんな思ってきたところとして、DNAからRNAができて、たんぱく質ができて、たんぱく質が最後は酵素としていろいろな分子を代謝してという基本的なゲノム、トランスクリプトーム、プロテオーム、メタボロームというオミックス階層を調べていくことで、ゲノムの結果から分かる遺伝子のターゲットを見つけたい。そういったときに、例えばプロテオームを狙ったときに何がいいかということ、現在でも薬剤標的の多くはたんぱく質ですので、病気の遺伝子の変化によってどのようなたんぱく質が動くかということが原因となるたんぱく質を見つけることにつながって、薬を見つけていけるのではないかとということになります。これが目的としてははっきりと一つの道筋として考えています。

ですが、それだけではなくて、生体のビッグデータを取るといろいろなことが分かってくるということで、特に次のところですが、ゲノムデータは生まれつきのデータですので、生まれつきの遺伝的リスクを求めていって、そこから生まれつきどういった方向になりやすいかが分かるということですが、その途中の人生の過程での時系列のオミックスデータを取っていると、さらにそこからどう変化していくことでどう変わっていくかといったことが詳細に分かってくることで、徐々に健康と疾患との境界線がどこにあるのかを生まれつきのゲノムデータ、さらに時系列のオミックスデータを1人の人についてどんどんたくさん取っていき、さらにそれを大人数集めていくと。それを計測しながら解析、開発していき、解析しながらまた新たな計測技術を開発していくということで、新たにいろいろなことが解明されていく。創薬が分かったり、病態が分かったり、予測ができたりということにつながっていくでしょうと考えているということで、こういったシンプルな理解からたくさんのデータを取っていくということが、まず基本として我々は用意しておくべきなのではないかと考えています。

最後に、紹介しろと言われたのでバイオバンク・ジャパンの紹介をさせていただきますと、今、B-cureという枠組みの中で我々はやっています、東北メディカル・メガバンク、ナショナルセンター・バイオバンクネットワーク、バイオバンク・ジャパンということでデータを整備しております。これらを統合してデータを出していきたい。先ほど岡田先生がおっしゃったように、データをどこかに置いておくのではなくて、これらのバイオバンクから出たデータをさらに国内で幅広く共有していきたいといったところで頑張っているということになります。

バイオバンク・ジャパン単体について御説明しますと、次のページに書いてありますけ

れども、2003年から収集を始めまして、2018年の時点で一旦断絶があったのですけれども、基本的には連続的に臨床情報をずっと集めているということと、ゲノムデータを全員集めているということと、血清のデータを2003年から2013年まで集めておりました。

疾患バイオバンクがよかったと考えておりますけれども、次のところにはBBJの研究成果ということで多数の成果が出ています。ここについて日本でもいろいろお話ししていて、日本でデータを取るというのが本当に成果につながるのか、本当に日本で、今、欧米であれほど進んでいるのに今さら取る必要があるのかと。ちゃんと取って、ちゃんと成果が出ると考えています。さらに、国際的戦略をちゃんと取らなければ埋もれますので、国際共同研究もちゃんとやっていく必要がありますけれども、国内でもちゃんと基盤を築いてしっかりとデータを取っていったって解析をしていくことの上に国際的な発言力が出てくるものと思いますので、その両者、お互いに併せて頑張っていけばいいなと思っているところです。

一番最後ですけれども、ここまで言ったようなことを実現するためにどういうことを私たちはやろうとしているのかというと、臨床情報と全ゲノムシーケンス、プロテオーム、メタボローム、ここの予算の出どころと相談しまして、抗体プロファイルなどを取って大規模なゲノム・オミックスデータセットをつくるということですが、それを単一の計算機環境に集めて、それに幅広いたくさんの研究者の方々にアクセスしていただくといった環境をつくっております。これは我々の努力として、まずはこれで何とか成果を出して、これはうまくいくものだ、こういったことは良いことだということでぜひ分かっていたいて、さらに広く大きくもっと大規模なデータ、もっとたくさんの研究者の方に集まっていたくように続いていけばいいなと思っているところになります。

私からは以上になります。

○長野次長 ありがとうございます。

米村構成員、最初のところは入っておられませんでしたけれども、既に入られたということで御連絡をいただいておりますので、共有させていただきます。

続きまして、AI活用の可能性という観点で、Preferred Networksの小林様よりどうぞよろしく願いいたします。

○小林参考人 よろしく願いいたします。改めまして、このような貴重な機会をいただきまして、ありがとうございます。

せっかくの機会でございますので、AI開発企業からというところで、弊社Preferred Networksの御紹介を少々2枚目にさせていただいております。大きく3つございまして、1つは国内でもトップクラスのエンジニアを取りそろえているという自負をしております。今、ざっと200名ほどのエンジニアを弊社では有しております。2番目がそういったAIエンジニアが実際に大量の演算を処理していくための計算資源として、スーパーコンピューターを自前でつくっております。こちら電力効率が世界ナンバー1を3回取っているといったスーパーコンピューターを有しております。3番目が様々な企業様とのいわゆるB to B

という形でのビジネスモデルを進めておりまして、こういった人的資源あるいは計算資源を活用しながら作業仮説の構築から行っていくという形で、今、国内最大のいわゆるユニコーン企業という形に成長を遂げているという企業でございます。

続きまして、3枚目に我々の注力事業を4つ挙げておりますけれども、その中でもライフサイエンスあるいはマテリアルサイエンスというのは注力事業の大きな一つになっております。中でもゲノム医療に関しても非常に大きく注目をしているところでございます。

4枚目にAI開発の観点からゲノムデータベース、これの基盤活用に対して幾つか整理をさせていただいております。まず、AI企業としてどのような貢献ができるかというところですが、一つは、当然ながら将来の医療への応用を中長期的に目指していくような医学的な貢献があらうかと思っております。もう一つは、短期的に現時点でこういったAI技術が必要としているような患者様、御家族あるいは医療者に対しての医療的な貢献もあらうかと思っております。この整理の下に幾つか具体的なユースケース、ほかの一般的な事例を含んでおりますけれども、そちらを御紹介しながら具体的なユースケースの共有をさせていただければと思っております。

続きまして、5枚目に遺伝子変異の病原性、いわゆる疾患を引き起こす特性を進化に伴って変異するような配列分布をモデル化することによって連続的に定量化したという事例を御紹介しております。こちらは遺伝子変化から生じる疾病リスクというものを進化というもの、要は生物学的な時間を、進化的な時間を考慮してモデル化した事例でございます。このような病原性を今まで不明だったものを定量的に評価できるということで、疾病リスクをより正確に捉えることが可能になっているという事例でございます。

2つ目が6ページにありますようなゲノムドリブンの創薬プロセスに機械学習がどのように活用されているかという事例でございます。こちらがいわゆる従来の遺伝子解析ではなかなか治療標的としての意義が見いだせなかったようなリスク遺伝子に関して、より機械学習を活用することによって高精度のリスク遺伝子を同定して、この事例ではアルツハイマー型認知症ですが、新たな治療標的あるいは創薬ターゲットを発見してきたという事例でございます。

続きまして、スライド7枚目が診断における精度あるいは効率性の向上といった事例でございますけれども、こちらは全ゲノム配列から大規模なゲノムデータベースを参照し、さらにAIアルゴリズムを使うことによって高精度かつ迅速な診断を可能にした事例でございます。これは遺伝性疾患を抱えて生まれてくるような新生児を対象にした研究でございますけれども、疾患の原因となるような遺伝子エラーの上位2種類いずれかを92%の精度で迅速に特定することができたという事例でして、こういった希少バリエーションに関する十分な情報を随時参照することができる。臨床現場でもこういったAIアルゴリズムを活用することによって迅速かつ正確な診断が可能になるのではないかと考えられています。

8枚目が治療予後に関する予測精度、これをゲノムと機械学習によって向上させていく事例になります。こちらは胃がんをターゲットにした、今、様々な治療モダリティがご

ございますけれども、それに対しての治療反応性を事前に予測していくような研究でございます。これも独自の機械学習アルゴリズムを使って遺伝子発現シグネチャーのパターンから治療の予後反応性を予測していくということで、こういった治療転帰の情報を備えましたゲノムシーケンスデータがあれば、より正確に予後予測を効率的に行うことができるのではないかと我々としても考えているところでございます。

そういった具体的なユースケースを基に、現状の課題からAI開発企業の立場で少し御提案をさせていただいております。ページをおめくりいただいて、10枚目にユーザビリティの観点からゲノムデータベースに対する御提案を少々させていただいております。

1つ目が、まずサンプルの多様性というものが一定レベル担保されていることが弊社のようなAI開発企業にとっては重要でございます。理由としては、データの偏りというのがモデルの精度あるいは般化可能性に大きく影響を与えるということがございます。ですから、例えば年齢、性別あるいは社会経済学的な因子と書きましたような教育歴あるいは家族歴、そういったところも含めた多様性がある程度担保されていることがモデルの構築には必要と考えております。

もう一つは、日本人のゲノムデータベースの構築において、どの遺伝子変異が人種特有のものであるかが可視化されていると、モデル開発においても人種非依存的なモデルを開発することができるのではないかと考えております。

もう一点、ユーザビリティの観点からは、データサイエンティストあるいはAIエンジニアに与えられた多くのタスクというものはある程度ターゲットが明確でございますので、検索性が高い、すなわちメタデータの検索において非常にその辺りがいわゆるユーザーインターフェースで実装されているということが我々としては比較的重要かと考えております。

最後に、現状のゲノムデータベース、幾つか我々も活用させていただいておりますけれども、なかなかどう使っていか分からないという声がエンジニアからは聞かれておりますので、そういった方法がチュートリアルみたいなもので分かりやすく示されていると活用しやすいのではないかと考えております。

最後のスライドで11枚目でございますけれども、ゲノムデータベース活用に対しての少々御提案をさせていただいております。我々AI開発企業としては、バイオバンクとの共同研究あるいは製薬企業とのいろいろなプロジェクトを進めて参りましたがけれども、なかなか基盤開発を進めるに当たっては実用化の目的が不明瞭であったり、或いは創薬のプロジェクトでもゲノムデータの活用はあまりなされていなかった。結果としては従来の疾患メカニズムに依拠したような分子生成モデルというところにとどまっていたということがございます。そういう意味では、そういった異業種、他業種の連携をもっとより明確に進めまして、業種を超えたようなタスクフォースを組成して、今、先生方におっしゃっていただいているようなゲノムドリブンの創薬であるとか、様々なプロジェクトを実験的に進めるような異業種タスクフォースというものを提案したいと考えております。そういった

ものをある程度逆にオープンに見える化して、チュートリアル或いはアーカイブという形で残すことで、同様の取組を様々な業種の方がしていくことも必要ではないかと思えます。

ありがとうございます。以上でございます。

○長野次長 ありがとうございます。

続きまして、主にヘルスケア領域における利活用といった観点から、NTTデータの田端様より御発表をいただきます。よろしくお願ひいたします。

○田端参考人 本日はお忙しい中をお集まりいただき、ありがとうございます。NTTデータの田端佑介と申します。

私自身、東北大学に出向してゲノム疾患リスク予測の研究をしている真っ最中でございまして、本日は研究者としての目線とビジネスを行うビジネスマンとしての目線、この両面から現在の業界の状態とCANNDsへの期待ということでプレゼンを始めさせていただきます。どうぞよろしくお願ひいたします。

それでは、2ページとなります。近年、ゲノム解析業界ではゲノムデータが蓄積されることで研究が進み、徐々にゲノムの活用事例が出始め、ゲノムへの理解が浸透しつつあります。ビジネス化に向けては大規模なデータを解析できるアプリケーションと基盤の開発がセットとなりますが、IT化の遅れが業界課題と言われております。

3ページです。NTTデータは参入障壁であるゲノム解析知識の習得から立ち上げ、サービスを提供しながらコア技術を獲得し、プラットフォーマーの中核企業として社会インフラの構築を目指しております。特に大学との共同研究を中心とした技術提供と検証で、ビジネスフィージビリティ調査を行い、コア技術を育てています。

4ページになります。例えば東北大学とは、ゲノム疾患リスク予測に取り組んでいます。東北大学の機械学習技術をコア技術とし、共同開発及び大規模データでの検証を実施しています。私自身も東北大学に出向してございまして、生のデータを触って解析をするとともに、会社に知見を伝えています。この取組では、糖尿病、痛風、高脂血症といった生活習慣病を対象に、従来の10倍以上の被験者数、10倍以上のバリエーションで検証を行っております。その結果、現状の各手法の精度と基盤に求められる計算量を確認することができました。この取組はビジネス化に向けて研究を実装し、ゲノムデータ解析サービスの開発・技術提供を2023年度から本格化させる予定です。

5ページになります。このような取組を通じて様々なサービスが提供可能になると考えています。医療ではゲノムデータの解析結果と各種医療データを組み合わせて、幅広いサービス提供が可能です。冒頭、岡田先生や鎌谷先生からゲノム創薬のお話もありましたが、そちらももちろんです。従来のゲノム創薬や疾患リスク予測と比べて、ゲノムは環境の影響を受け難く、既存の解析に疾患の遺伝的要因を加えることができるのがメリットだと考えています。

6ページです。医療以外にも様々なサービスを提供可能です。例えば情報銀行、アグリゲノミクス、ゲノムを用いた新商品の開発、予測の高度化などが挙げられます。医療以外

では今後新たなサービス開拓が行われる可能性も高いのではないかと考えています。

7 ページです。ただ、その一方で、ビジネス化を阻害する主な課題も見えてきました。1 つ目の課題は、ビジネスの前提にあるゲノムデータが入手しづらいということです。ゲノムデータは究極の個人情報と呼ばれます。そのため、各コホートの審査は厳重で、他社からの融通も利きません。また、自社でデータ収集をするとすると1 検体当たり数万円かかるため、人数を集めると大変高額で、1 社では厳しいものがあります。さらに、各データは収集目的・商用利用が限られており、取得地域や年齢層、性別などに偏りが見られます。2 つ目は、投資判断がつかないという課題です。市中の情報が少なく、かつ専門家向きで、ビジネスに寄与できるレベルの手法や精度、基盤環境が見えていません。そのため、簡単な検証でもして確かめたいところなのですが、その検証自体も容易ではありません。さらに、ゲノム解析技術やAI技術を保有する人材が不足しているという課題もあります。ただ、これらの課題に関しては、国が主体で進めるCANNDsが実現すれば解決できる可能性が高いのではないかと考えています。

8 ページになります。CANNDsとは図のような連携をイメージしています。ビジネス化で課題となる大規模なゲノムデータはCANNDsを活用し、弊社はゲノム解析サービスに注力したいと考えています。また、投資判断となる検証や人材育成にもぜひCANNDsを利用したいと考えています。

9 ページになります。最後にCANNDsへの要望を述べます。

1 つ目の要望は、大規模なゲノムデータと分析基盤を提供して欲しいです。小規模なゲノムデータは自社で収集するとしても、大規模なゲノムデータや大容量のメモリを要する検証基盤はCANNDsを利用したいです。また、ある程度のデータクレンジングや計算はCANNDs側で実施し、ゲノムデータに反映しておいて欲しいです。

2 つ目は、各種調査で集めた調査票を公開して欲しいと思っています。最低限、参加時年齢、性別、健康診断の値などの形質値、データの品質に関する情報や解析手段は公開して欲しいです。例えば私が研究に従事していますゲノム疾患リスク予測における東北大学の機械学習手法の場合、各被験者個人のバリエーションが入っているゲノムデータが欲しいです。被験者数は数十万人規模、バリエーション数は数千万バリエーション規模、メモリは数百ギガバイト、データ形式はバリエーションの情報が入っているPLINKで扱える形式のデータ、形質として健康診断の値、データにはゲノムデータの品質に関する情報が付与されて欲しいです。

10 ページになります。続いて、セキュリティーを意識しないレベルのゲノムデータのハンドリングをお願いします。セキュリティーの確保はできる限りCANNDsで実施して欲しいです。

次に、解析結果は幅広く商用利用可能であって欲しいです。商用利用に関する同意書のようなものがあれば良いと考えています。

最後に、CANNDs側やほかの研究で実施した解析結果を公開して欲しいです。商用利用や利用者側が解析の正しさを確認できるため、最低限GWASの要約統計量などの解析結果を公



開して欲しいと考えています。また、審査などはあると思いますが、個人レベルの解析結果が利用できるとなお良いと考えています。特に論文や特許などの権利が切れたものは積極的に出して欲しいと思います。公開場所に関しては特に希望はございません。

発表は終わります。御清聴ありがとうございました。

○長野次長 田端様、かなり具体的な要望も含めて御発表をいただきました。ありがとうございました。

それでは、最後になりますけれども、ゲノム解析の技術開発といった側面から、Mitate Zepto Technicaの橋本様より御発表いただきます。どうぞよろしく願いいたします。

○橋本参考人 ただいま御紹介をいただきましたMitate Zepto Technicaの橋本と申します。本日は貴重なお時間をいただきまして、誠にありがとうございました。

私どもMitate Zepto Technicaは、ゲノム医療におけるアクセス、クオリティー、コストを改善するために、半導体技術を活用したハードウェアソリューションを提供するという目的で2020年に設立されましたベンチャー企業でございます。現在はゲノム解析の中でも課題となっておりますシーケンス後の解析、こちらの解析時間を大幅に短縮するためのゲノム解析専用の半導体の開発を行っております。また、それを搭載したアクセラレーターボードの開発を行っておりまして、来年度中には市販化したいということで奮闘しております。

半導体のASICというのは、特定の機器や用途のために専用に設計、製造される半導体のことをいまして、その用途に特化している分処理スピードが速くなり、かつ無駄な能力を逆に言うと削ることができますので、コストも下げることができるというものになっております。

現在開発しておりますアクセラレーターボードは、通常のワークステーションに搭載を前提で設計をしておりまして、現在私どもがお伺いしている限りですと、全ゲノム解析で主流になっておりますのはスーパーコンピューターを使われることが多いということでございますので、こちらはスーパーコンピューターを使わず通常のワークステーションを使って全ゲノム解析を行うことができるようにする予定で動いております。ちなみに、スーパーコンピューターの消費電力に関してはページの14のほうですね。後ろから4ページのところにつけておりますので、御参照いただければと思います。通常のワークステーションを使うことでかなり大幅に消費電力を下げることができますので、専用の電源を確保する必要もないということで、一般的な医療施設でも導入することが簡単になるのではないかと考えております。

具体的にどのような形で効率を上げるかというところになるのですが、シーケンス後のデータ解析、現在大体30分とか50分かかると言われておりますが、私どものハードウェアを使えば5分で解析できるようになるのではないかと考えております。こちらは1日10人から100人という形で書いておりますが、私どもは研究というよりは実際に医療機関で使っていただくことを想定しているのです、1日を8時間の時間で計算した状態になっており

ます。従って研究施設で24時間使うとなれば、単純にこの3倍の数字で見ただけであればいいかと思えます。ですから、このような形で効率が大幅にアップされますので、通常の医療機関においても解析機器を導入することが容易になるのではないかと思っております。また、導入コストも先ほど申し上げましたASICという半導体を使えば、現状主流になっておりますアクセラレーターボードの約10分の1の価格で抑えることができるかと思っておりますので、導入コストも10分の1になる予定でございます。

次のページになります。ヒトゲノム解析のコスト、こちらの表はいつも出てくる表ですので、皆さんよく御覧になっているものだと思います。アメリカの国立衛生研究所の出したグラフになりますが、こちらの白い線、ムーアの法則と呼ばれているものがございしますが、こちらは半導体業界で言われている指標になります。1年半から2年で半導体の性能は倍になると言われておまして、同様にコストも半分になると。すなわち、そのぐらい速いスピードで半導体業界は進歩しているのですよというものを表している表なのですが、グリーンのラインで示しますとおり、ゲノム解析コストはそれを超えて一気にコストが下がっているというものを示しているものでございます。しかしながら、2005年の次世代シーケンサーの登場がこのムーアの法則を超えたコスト削減の切っ掛けになっているのですが、2010年ぐらいでそれも落ち着きまして、それ以降はまた同じくムーアの法則のスピードにのっかって解析コストが下がっている状況になっております。ですから、現在私どもが計画しておりますこのアクセラレーターボードが実現すれば、次世代シーケンサー同様のインパクトをこの解析マーケットに与えられるのではないかと考えておるところでございます。

続きまして、次のページになります。こちらも皆さんよく御存じのとおり、2019年からがんの遺伝子パネル検査が保険適用になったわけですがけれども、しかし、実際に受けられる患者の方はまだまだ少なくて、標準治療がないもしくはそれが終わった患者様のみが受けられるという状態でございます。すなわち、まだゲノム医療というものは身近な医療ではないと私どもは考えております。そうであるからこそ、私どもとしてはゲノム解析をより身近にするためにこのアクセラレーターボードを考えているというところでございます。

次のページになります。計算機のリソースの適切配分という形で表現させていただいております。私どものアクセラレーターボードが実現すれば、ワークステーションでシーケンス、解析ができるようになりますので、医療機関にシーケンス、それから、解析まで任せることができるようになります。そうすれば、CANNDsまたは産業界・アカデミアの方はスーパーコンピュータをより高度な解析に使うことができますので、実際の全ゲノム解析のところは医療機関で終わらせるということを目指したいと考えております。

次のページになります。参考資料1の「全ゲノム解析等実行計画2022」のページで行くところの32ページの資料を使って改定したものになるのですが、私どものアクセラレーターボードが実現しまして、実際に医療機関においてシーケンスから解析までができるようになれば、医療機関からCANNDsに直接データを送ることもできますし、当然そこに対す

る患者様へのデータを戻すというところのスピードも大幅にアップすることが言えるかと思えます。

次のページも同様でございます。つまり、現状、間にシーケンス企業様が入っているかと思いますが、こちらをなくして医療機関で検査をすると。ただし、もちろんシーケンス企業様は大変重要なポジションにございまして、例えば未知のゲノム解析ですとか、そういったものに関しては引き続きシーケンス企業様にやっていただくということになります。

こちらはなかなか時系列的に説明が難しいところにはなるのですが、次のページ11になります。シーケンス企業様もそうなのですが、CANNDsの方にある程度大量のデータが溜まってくれば、少しずつ先ほど岡田先生や鎌谷先生がお話しされたような世界が実現する訳ですが、まだそこに行くには時系列サンプルのデータであったりですとか、大量のシーケンスデータを取っていかなくてはいけないということになっているかと思えますので、そこに私どもとしては寄与していきたいと考えているところでございます。将来に関しては、個別化医療、個別化予防に向けて動いているところだと思えますが、その基になる大量のゲノムデータがまだ産出できていないところでございますので、私どもとしてはここに技術を投入していきたいと考えておるところです。

しかしながら、そこが実現したときに出てくる課題としましては、例えば医療機関によって診療データですね。そういったものをどう管理していくのかですとか、民間事業者で既に各種ライフログデータが取られております。例えばウェアラブルデバイスを使って取った計測データほか、様々なデータが既に世の中には取られているわけですがけれども、そこをどのように管理するかが必要になってくるかと思えます。また、そのような大量のデータが今度できたときに、果たしてそのデータを扱う人材が足りているのかというところも問題になってくるかと思えます。さらに加えて、こちらも皆様方からの御提案にもありましたとおり、国際的な標準化を行っておかないとなかなかそこについていけないのではないかと考えております。

最後に、私どもからの要望となりますが、私どものような医療現場を全く知らずに入ってくるハードウェアのソリューションを考えているような事業者にとって、開発の中でゲノムのデータというのは、正直なところ、ネットで探せば割と出てくるころではあるのですがけれども、やはり倫理的や法的、社会的に使用すべきデータではないと認識しております。また、再現性の高い技術検証をするためには、出自の確かな精度の高いデータを必要としております。なかなか新規参入ですと業界とのネットワークが乏しく、適切な期間、タイミングでこのようなデータを取得するのが困難となっております。ですから、ぜひCANNDsを通じて私どものような新規参入事業者にとっても扱いやすいというか、アプローチしやすい形でデータを提供していただくと大変助かるかと思っております。

その中では当然倫理的な問題であったり、法的な問題があるかと思えますので、提案としましては、Precision FDAなどがChallegeで様々な課題を提案型イベントとして出されて

おりますので、そのような形でデータを出していただければ、私どものような医療業界とは違うところからその課題を見て、自らの技術で対応できるのではないかと、全く違う業界からこの課題を改定するアプローチがでてくるのではないかと考えておりますので、ぜひそのような形で解決型のイベント等を開催していただけると、すごく私どもも現状この業界において置かれている問題等に気づきやすいのかと考えておるところでございます。

以上となります。ありがとうございました。

○長野次長 ありがとうございました。

それでは、5名の先生方からそれぞれのお立場で御発表いただいたところですが、その中で、まさにゲノム解析の関連技術が日進月歩であること、それから、利活用のされ方についてもかなり多様化されていると。そういった中で現状の課題、それから、今後に関する御要望、御意見などをいただいたところかと思えます。

それでは、時間があまりございませんが、5分程度になりますけれども、御質問、御意見などがございましたらよろしくお願ひいたします。

○上野構成員 製薬協の上野でございます。

いろいろなお話を聞かせていただいて非常に勉強になったのですが、最初の岡田先生のお話の中で、海外ではゲノム情報のコンソーシアムの形成によって広くそれが進んでいるとか、バイオバンクもそういう流れかと思うのですけれども、こういう横に繋げるときに一つの大きな課題が標準化というところがあるかと思うのですけれども、それぞれ全て同じ条件でやられているものではないものをつなぐときの標準化という点をどのように乗り越えてそういうコンソーシアムを設けて共有をしているか。そういう点についても何か情報がありましたら御提供いただければと思います。

○岡田参考人 恐らく作成するデータの標準化についての御質問かと思われま。いわゆる新たにデータをつくる場合はワーキンググループがデータ解析、サンプル収集、幾つか立ちまして、相当綿密に議論した上で、これならば安全かつコストパフォーマンスよく標準化されるとプロトコルが確立した段階でプロジェクトがプロポーズされるといった状況を取っております。ですから、そこまでできる担保ができていないとなかなかコンソーシアムというものが動かないし、予算もつかない、そういう体制かと感じております。

○上野構成員 ありがとうございます。

○長野次長 上野構成員からの御質問に岡田先生から御回答いただきました。ありがとうございます。

中釜参与、挙手されていますでしょうか。

○中釜参与 中釜です。御発表ありがとうございました。

御発表の中で、Mitateさんの資料についてお伺いしたいことがあります。計算技術の進歩は非常に素晴らしいと拝聴させていただきました。今、進められている全ゲノム解析事業において、医療機関の中でシーケンス解析をして、それをレポート作成し、患者還元するという仕組みができれば非常に良いと思うのですけれども、現在全ゲノム解析のデー

タを患者還元するというスキーム、結果の妥当性等々についてまだ十分には検討し切れていないと思います。その辺りはこのスキームの中で今後どのようにしていかれるのが良いとお考えなのでしょう。その点を教えていただければと思います。

○橋本参考人 ありがとうございます。

私どもはハードウェアの方の設計を中心にしているものですので、具体的に倫理的な部分であったりですとか、法的な部分、かなり議論されていることは承知しておりますが、まだまだそういった部分では浅い部分もございますので、私どもとしてはこういうことまでできますよというところまでしか御提案できないかと考えております。申し訳ございません。

○中釜参与 分かりました。ぜひ技術の革新とともに患者さんに還元する際の感度、特異度の問題、あるいは結果の妥当性、臨床的有用性等について、こういうシステムの中で技術開発とともに一緒に議論していくようなスキームを御検討いただければと思いますので、コメントさせていただきました。ありがとうございます。

○橋本参考人 ありがとうございます。

○長野次長 他にはよろしいでしょうか。また先に進めさせていただいて、後ほど戻られても結構でございます。

せっかく5人の専門家の方から御発表いただきましたので、簡単に私のほうで現時点でまとめさせていただきますと、様々な具体的な御要望も含めて提示いただいたところですが、その中で割と共通的にもされたこととして5つあるかと理解いたしました。

1点目として、例えば年齢や地域、他にもございますけれども、属性に関してデータの多様性というものが重要であること。

2点目で、大規模なゲノムデータをいろいろな方々が、敷居が高くなく利活用できる環境が必要であること。

3点目に、ゲノムデータそのものも重要でございますけれども、それに加えて付随するような健康・医療情報ですとか、オミックスデータですとか、そういった利用可能なデータを増やすことによってゲノム研究の可能性を拡大することが重要であること。こういったことは各ゲノムやバイオバンクなどでも取り組まれておりますし、さらにCANNDsそのものでもそういったことを目指しているということで、目指すべき姿であると思っております。

また、もう一つがデータの共有や公開の在り方ということで、GWASのような解析結果もしっかり公開するという一方で、さらに研究開発に利活用できると。例えばGWASのインテンシティーデータですとか、元データに近いようなデータもしっかり公開するといったことも重要だろうと理解しました。

最後になりますけれども、5点目としては、異業種あるいは異分野の専門家の方、色々な方々同士の協力の重要性ということについても理解したところでございます。

他にも具体にも色々ございますけれども、割と共通的な部分について述べさせていただきます。

きました。次に、議題2)の「全ゲノム解析等実行計画の進捗について」ということで進めさせていただきます。厚生労働省よりその進捗についての御説明をよろしくお願いいたします。

○城審議官 厚生労働省大臣官房医薬産業振興・医療情報審議官の城でございます。

全ゲノム解析等実行計画の進捗について御報告をさせていただきます。よろしく申し上げます。

資料は資料2と記載されているものになります。

2ページを御覧ください。こちらは前回説明したものと同様になりますが、全ゲノム解析等実行計画の推進に関する政府方針等を記載したものでございます。経済財政運営と改革の基本方針2022等におきまして、臨床情報と全ゲノム解析等の情報を連携させ搭載する情報基盤を構築し、その利活用に係る環境を早急に整備する、新たな治療法を患者さんに届ける取組を推進する等の文言が掲載をされているところでございます。引き続き厚生労働省としても本実行計画を着実に推進していきたいと考えております。

次、3ページを御覧ください。令和4年度がん領域のAMED研究班の概要についてお示しをしております。がん領域につきましては、この表に記載しておりますとおりA班、B班、C班、それぞれA班3班、B班6班、C班1班の構成となっております。本年度はA班ではそれぞれ中の各班で各600症例ずつの全ゲノム解析を行って患者還元を推進しています。B班につきましては、昨年度解析したデータにオミックス解析等を追加いたしまして、多層的なデータ解析に基づく研究を行うこととしております。また、C班につきましては、一番下でございますが、集中管理チーム、ゲノム解析チーム、臨床情報チーム、レポート作成チーム、データ共有チーム、出口戦略チームの6チームに分かれて研究を進めているところでございます。

4ページを御覧ください。がん領域のAMED研究班におけるこれまでの成果をお示ししております。これは令和4年の10月20日の時点でございますが、全ゲノム解析等の結果に基づくエキスパートパネルは677症例で実施をされておりまして、そのうち治療薬の選択やがん種の診断、遺伝性疾患の診断に有用な病的意義がある変異の検出が381症例、およそ56%になりますが、ここで認められたということでございます。また、既存の検査では検出できないがんに関するゲノム異常の検出が113症例、真ん中の全ゲノム解析の結果が診断に有用であった症例が32症例、がん以外の疾患に関与する可能性が高いゲノム異常の検出については30症例で認められております。これは一部1と2、1と3で重複がございます。がんにつきましては、この重複を差し引いてになりますが、全ゲノム解析ならではの成果が得られたものは重複を差し引いて131症例でありまして、全体のおよそ20%となります。これらの成果につきましては、出口戦略の加速による創薬の促進、患者還元の拡大につながるものと考えられます。

5ページを御覧ください。全ゲノム解析等の実施により検出できたゲノム異常の具体例を2つお示ししております。上段でございます。上段は全ゲノム解析の結果が診断に有用

であってということで、全ゲノム解析の構造解析によりまして、組織型が不明であった頭頸部腫瘍におきまして、これは染色体の転座による多形質性副鼻腔肉腫で頻発することが知られているPAX3遺伝子とMAML3遺伝子の融合遺伝子が生じていることが検出をされたところであります。この融合遺伝子が検出されたことによりまして、正確な組織型の診断に至ることができたという例でございます。

下段でございますが、これは既存の検査では検出できないがんに関与するゲノム異常の検出ができた例となります。耳下腺腫瘍の症例におきまして、全ゲノムの構造解析によりまして、腺房細胞がんで頻発することが知られている染色体転座が検出されました。この転座はenhancer hijackingによるNR4A3遺伝子の発現上昇を誘導することが報告をされておりまして、本症例においても検出された染色体転座が本症例の発がんに関与するドライバー変異である可能性が示されたところであります。

6 ページを御覧ください。こちらは令和4年度の難病領域のAMED研究班の概要をお示ししております。難病領域の研究班につきましては、国立国際医療研究センターの國土先生が研究代表者を務める1班で構成をされております。体制は御覧のとおりでございます。

そのまま7 ページを御覧ください。これは國土班の令和2年度から3年度の解析状況でございます。令和3年度までの実績としましては、6,861検体について全ゲノム解析を実施いたしております。内訳は御覧のとおりでございます。この中の診断困難な疾患（未診断疾患）のうち、全エクソン解析でも疾患の特定に至らない患者、これは患者数であります。809症例のうち76症例、9.4%が全ゲノム解析を通じて疾患の特定に至るとということが示されたところでございます。

8 ページを御覧ください。これは全ゲノム解析等実行計画2022の概要でございます。これまでの検討を踏まえまして、全ゲノム解析等実行計画2022が令和4年9月に策定をされております。こちらに概要についてお示しをしておりますところでございます。全体版につきましては参考資料がございます。先ほど触れておられましたが、御参照いただければと思います。

全ゲノム解析等実行計画の第1版、下のところの左側になりますが、これが策定されました令和元年度から令和3年度までにつきましては、既存の検体を用いた先行解析として約1万9200症例の解析を行って、本格解析の方針決定と体制整備を進めてきたところでございます。令和4年度からは新規の患者の検体を用いた本格解析を開始いたしております。本格解析におきましては、今回新たに策定されました全ゲノム解析等実行計画2022を踏まえまして、戦略的なデータの蓄積、解析結果の日常診療への早期導入、新たな個別化医療の実現を通じて、国民へ質の高い医療を届けることを目標としております。表の上から3列目の「解析実績・予定」のところに記載をしておりますが、令和4年度以降は10万ゲノム規模を目指した解析の他に、マルチ・オミックス解析等を予定しておりまして、また、患者還元体制の構築や情報基盤や事業実施組織発足に向けた検討、ELSI・PPIに係る事項等についての検討を進めていくとしております。

次の9ページを御覧ください。全ゲノム解析等の推進によって目指す医療の姿を図解したものでございます。解析結果の日常診療への導入、新たな個別化医療の実現、蓄積されたデータを用いた研究・創薬を推進して、患者への還元や質の高い情報基盤の構築を行い、これらの取組について国民へ質の高い医療を届け、がん・難病等の克服を目指すことといたしております。

10ページを御覧ください。これは全ゲノム解析等実行計画の目次をお示ししておりますが、この中で先ほどお話のありました7、8、倫理的・法的・社会的課題に係る事項、それから、患者・市民参画に係る事項について11ページに抜粋をしております。

11ページを御覧いただければと思います。この7のところはELSIに関する事項でございます。ここは事業実施組織にELSI部門を設置して検討、対応を行うこととしております。特に留意が必要な事項として、下の方でございますが、ゲノム情報等による不利益が生じないようにするために社会環境の整備をする上での制度面の課題を把握した上で、本事業における対応方針を検討することといたしております。

また、8のPPIに係る事項でございます。事業実施組織に患者・市民参画部門を設置いたしまして、本計画に参画する研究機関・医療機関においても患者・市民の視点を取り入れるための体制を設けることをいたしていきたいと考えております。それによりまして、広く国民向けの情報発信・周知活動を実施するとともに、患者・市民からの意見を集約し事業に反映させるための体制を構築することといたしております。

12ページを御覧ください。こちらは事業実施組織の発足に向けた中長期的なスケジュールをお示ししております。上段の「患者還元」でございますが、これは今後も全ゲノム解析の検査としての有効性の確立に関する検討を継続することとともに、出口を見据えた臨床試験・治験等を実施、推進をしていくこととしております。それから、中段の「データ基盤としてのインフラ整備」でございます。これはシステム構築に着手していくということとともに、解析データ・臨床情報の蓄積を並行して継続をし、段階的なデータ提供による利活用の開始を目指していきたいと考えております。下段の「法人」でございます。これは本年度中に事業実施準備室を発足して、民間とも協力をしながら事業実施組織発足に向けた検討を進めるということといたしております。また、最適な法人の在り方についても検討を行って、今後必要に応じて法改正を行うなどの対応も行っていきたいと考えております。

13ページを御覧ください。こちらに全ゲノム解析等実行計画に係る事業実施体制の案についてお示しをしております。令和7年度中には厚生労働省が主催する全ゲノム解析等の推進に関する専門委員会の方針を踏まえまして事業実施組織を設立すること、これは予定でございますが、そうしたいと考えております。この事業実施組織の中にはボード、諮問委員会、利活用審査委員会、その他事業実施に必要な部門を備えておくと。それから、医療機関、シークエンス企業、解析・データセンター、産業界・アカデミア、この管理であったり、支援であったりということをしていく体制を構築していきたいと考えております。



また、実行計画にも記載されておりますが、解析・データセンターにつきましては、データ利活用ポリシー及びデータ共有ルールに基づきまして、産業界・アカデミアが研究・創薬のためのデータ利活用ができるようにということで、データ共有システムを構築し、データ利活用を推進するということをする予定にしております。そのために、事業実施準備室におきまして、令和4年度中にデータ利活用ポリシーやデータ共有ルールの策定、事業実施組織内のデータ利活用審査委員会の設置に向けて検討を行って、パイロット的な運用を行っていききたい、令和5年度中には本格的なデータ共有の開始を目指していききたいと考えております。

14ページを御覧ください。こちら、細かい資料になりますが、全ゲノム解析等実行計画に係る事業実施組織の行う事業の概要の案をお示ししております。左側の事業背景、目的、戦略等につきましては、計画に記載したものを抜粋しております。それに基づきまして、右側でございますが、事業内容を設定しております。具体的な事業内容としましては、この1)、2)、3)とずっとありますが、1)全ゲノム解析等の結果及び成果の速やかな患者還元の支援、2)として個別化医療の推進の支援、3)として質の高い情報基盤の構築と運用、4)として患者・市民参画の推進、国民向けの情報発信・周知活動の支援、5)としてELSIの支援、6)として人材育成支援などを想定しております。現在、研究班でこの事業実施組織の詳細な事業計画についての検討を進めているところでございます。

15ページを御覧ください。これは全ゲノム解析等実行計画に係る事業実施組織のビジョンを図示したものを示しております。上段に矢印が上を回ってきていますが、新たな個別化医療等の推進によりまして、これを日常診療に導入すると。これによって患者還元を行うと。そして、そこで得られたゲノム情報や臨床情報等、こういったものを戦略的なデータ蓄積を行うことによりまして、研究や創薬等への利活用促進を行うと。そして、情報基盤を活用した成果の創出を促して、それを患者還元につないでいくということで、国民に質の高い医療を届け、将来的ながん・難病等の克服を目指す好循環の実現を目指すとといったビジョンを持って検討を進めていききたいと考えております。

以上が厚労省におけます全ゲノム解析等実行計画の進捗についての報告となります。

○長野次長 ありがとうございます。

それでは、ただいまの厚生労働省の説明に関しまして、御質問、コメントなどがございましたらよろしくお願ひいたします。

それでは、山口構成員、よろしくお願ひいたします。

○山口構成員 お願いします。

私ども、今、御紹介があったA班の1グループとして行動していますので、全ゲノム解析の実行部隊の立場から今のご説明に少し補足させていただきます。既にがんに関しては、私どもで約2,000例の全ゲノム解析を終えていまして、そのうち一部がAMED関連の研究です。診断あるいは難病も含めて診断のレベルでは随分よくなっていると思います。先ほどの数字よりはがんの種類によって違うと思うのですが、私どものドライバー遺伝子の見つけ

方は、例えば肺がんですと、全ゲノムにより8～9割まで達しています。

お話がなかった患者にとって一番大切な薬物到達性、これもかなり全ゲノム解析の結果はいいのです。肺がんに限っていえば9割、肺がんを含む倍ぐらいの症例で5割ぐらいの薬剤到達性が認められているので、これまでのパネル検査に比べ、パネル検査は標準治療を終了した特殊な症例での検討ですが、多分3～4倍の薬剤到達性というデータが出ております。実際にそこまで来ると、現場として一番望んでいるのが新規薬剤です。ドライバー遺伝子は分かったけれども薬がないという症例が多く、そういう観点からゲノム創薬に大変期待をしています。

ただ、ハードルは非常に高く、この全ゲノム解析の結果のデータベースが出来上がったら日本の創薬が一気に進むかということ、そうではないと思うのです。理由は、この分野では欧米が先行していますし、既に創薬は始められており、また、欧米のメガファーマに比べて開発力は日本の場合まだ十分ではないと思います。過去には、日本の創薬が非常に輝いた時があって、抗生物質で醸造技術、お酒づくりの技術が大変役立ったという歴史もあると思うのですが、薬草や海洋産物、ドラッグリポジショニングあるいは化学合成といった創薬の歴史から見て、ゲノム創薬という観点からいうと日本独自の創薬資源、アイデア、さらには研究費の適正な配分、こういったことを意識しておかないと、日本発のゲノム創薬はまだまだ難渋するのではないかと考えています。

アイデアという点に関して一言だけ申し上げておきますと、今日のお話はみんな大きなデータベースをつくってというところなのですが、臨床医としては一人一人の患者の臨床データが非常に大事になってくると思うのです。私どもがずっとやってきたのは1万症例のがん患者の腫瘍の組織のデータを持っていて、それは全て電子カルテ上のデータと完全に紐づけることができます。そこから新しい知見も得られているので、その膨大なビッグデータとともに一例一例をじっくり見ていく姿勢も持っていく必要があるのではないかと思いました。

以上です。ありがとうございました。

○長野次長 ありがとうございます。

それでは、次に菅野構成員、よろしくお願いします。

○菅野構成員 菅野です。

資料の参考のところの18ページですか。解析予定数が新規患者で2,000人となっているのですけれども、数にこだわるわけではないのですが、10万人ゲノムを意識した計画で解析を進められるという御予定だったように思いますが、このペースだと10万人に達するのはかなり先になりそうな気がするのですけれども、これはそういう理解でよろしいのでしょうか。

○長野次長 厚生労働省、よろしくお願いします。

○城審議官 厚生労働省の城でございます。

目標は10万ゲノムであります。検体数とお考えいただければと思います。今回の実行計

画2022におきましては、がんと難病で10万検体分のゲノム解析を目指した解析結果のほか、マルチ・オミックス解析の結果等を含むということでやっておりまして、実際の対象症例数については徐々に延ばしていくことにございますので、毎年度、前年度の状況も踏まえて具体的に検討していきたいとは考えております。もちろん御指摘のようにまだまだこれからということは仰るとおりだと思いますし、それに向けて私どもはしっかりやっていきたいとは考えております。

○菅野構成員 そうすると、1人の患者さんから幾つものデータが出るという感覚で、それから、がんと難病と両方合わせてという感覚でという感じなのでしょうか。

○城審議官 大体難病だと物にもよりますけれどもトリオになりますし、がんの場合も最低2検体から比較していきますので、そういう意味では1人の患者さんというか1症例から複数検体が解析対象になると考えます。

○菅野構成員 妙に数にこだわってもしようがないと思うのですが、ぜひよろしくお願いたします。

○長野次長 それでは、次に上野構成員、よろしくお願いたします。

○上野構成員 最初の山口先生の御発言に対して、それを含めてコメントさせていただきたいと思うのですが、仰ったとおりそういう原因ゲノム、遺伝子が分かったとしても、それに基づく薬がない状況がある、そういう課題が十分あるということで、我々製薬企業としては、いかにそれに効く薬を提供できるかが大きな課題だと思っています。欧米の製薬企業に比べてという御発言もあったかと思うのですが、逆に我々日本の製薬企業でも規模によらずいいゲノム情報を基にコンパクトに創薬をすることができるのではないかと、そういう期待を持って我々も取り組んでいるわけで、だからこそ日本の質の高いゲノム情報を活用できるような環境をぜひ一緒になって整えて、我々もそれを活用する、そういうことを考えているわけです。

したがって、ゼロから創薬をするというのがありますが、これまで私どもがいろいろな研究で見いだしている既存の医薬品であり医薬候補品というものは我々の中で持ってござりまして、そういうものを新しいゲノム情報に活用できるということもあって、そういう点で考えると、今回産業フォーラムということもありましたけれども、こういった取組をぜひ加速させていただいて、こういったゲノム情報を基にした創薬、非臨床がどうあるべきかとか、或いはそういったものに出てきた候補化合物の臨床試験がどうあるべきかといった議論をこういった産業フォーラム或いはアカデミアフォーラムで議論していくことが重要だと思いますので、改めてそういったスピード感を持って議論できるような環境を整えていただきたいと思います。

以上です。

○長野次長 ありがとうございます。

厚生労働省から何かレスポンスはございますか。よろしいですか。

○城審議官 城でございます。

我々は実際にいろいろなルールを策定した上でパイロット運用などをしていくこととなりますので、令和5年度中にデータ共有の開始を目指す旨と記載しておりますように、しっかりと進めていきたいと思っておりますが、その際にはぜひしっかりと御協力をいただきたいと思っております。よろしく申し上げます。

○上野構成員 よろしく申し上げます。

○長野次長 他に何かございますでしょうか。よろしいでしょうか。

それでは、次の議題に移りたいと思っております。議題の3番目として健康・医療研究開発データ統合利活用プラットフォーム、CANNDsの関係の事業の進捗についてということで、このCANNDsは既存のバイオバンク、それから、今ほど厚生労働省から説明があった全ゲノム解析等実行計画における情報基盤、そういったものの連携基盤としてCANNDsを推進していくといった事業になってございますけれども、今年度が事業実施としての最初の年ということになります。このCANNDs事業の進捗について御報告いただきたいと思っておりますが、まずAMEDから御説明をよろしく願いいたします。

○松岡課長 ありがとうございます。日本医療研究開発機構データ利活用基盤整備推進準備室長、本日、小賀坂が御説明の予定でございましたけれども、小賀坂に代わりまして課長の松岡から御説明いたします。

資料を用いまして御説明いたしますので、まず2ページを御覧いただければと思います。CANNDsと書かれました二重線の四角がございます。こちらは健康・医療研究開発データ統合利活用プラットフォームを機能として書き下したものとなっております。上に「利用研究者（大学等）」「利用研究開発者（企業等）」とございますけれども、こういった様々な利用者にこの連携基盤を御利用いただくわけでございます。そこにはまずユーザー向けの統合的なユーザーインターフェースがあり、これを通じて利用者が様々な企業を利用することを想定しております。まずはこの四角の中の真ん中にございますようにメタデータを使いまして、この連携基盤を通し利用可能なデータの横断検索が行われることとなります。ミニマムな機能といたしまして、左にもございますように、IDの管理、ID連携機能、同意管理、認証／ログ管理といった基本的なものを搭載いたします。また、研究に資するようなプロジェクト管理でありますとか、結果閲覧といった機能も搭載予定でございます。

こちらの箱から下に向かって青い線で繋がっておりますのが、実際のデータへの接続でございます。左側の緑の四角が当面接続を目指して開発を進めております三大バイオバンク等のゲノムデータでございます。将来的には真ん中にございますように、厚生労働省様が進めておられる全ゲノム解析等実行計画のデータ、一番右にございますAMEDの研究開発から生み出されるデータ等が順次連結され、利用者に対して提供されているということを構想しております。

3ページを御覧いただければと思います。こちらはシステム面から見た事業の構想案をお示ししております。今年度の末を目途にまずはパイロット運用開始としておりますけれども、小規模での利用試験を開始いたします。2023年の年度中には限定的な提供からある

時点において広く一般利用者の提供へと移行することを予定しております。限定提供、一般提供、それぞれ右下の表のとおりに想定しております利用者、実施場所、実施内容等をお示ししているところがございます。なお、この左側の矢印に沿いまして、今後全ゲノム解析等実行計画との試行的な連携、それから、AMED支援研究開発データの提供というようにサービスを徐々に拡大していくということを考えております。

次の4ページを御覧いただければと思います。今後検討を進めていくべき論点の中から、他にも多々ございますけれども、重要なものをこのスライドにまとめております。まず1つ目の●でございますが、現在先ほど申し上げました三大バイオバンク等のデータや全ゲノム解析等実行計画のデータ、こうしたデータの提供者はストレージ及びデータ提供環境を整えておられます。連携基盤のほうはそれらのメタデータを収載することを前提としたシステム構想になっております。ところが、今後データの提供者のありようも多様化することが想定されます。例えばデータの提供環境は持たず、データしか持っておられないと。そうした場合も想定して今後のシステム開発は対応していく必要があるだろうということを考えているところです。

2つ目の●でございますが、データの利用者側におきましても多様化が予想されております。例えば御自身の持つソフトウェアや、あるいは連携基盤から利用可能以外の計算環境を利用したいといった利用者が存在した場合、どのようなシステムの構成にするのか。これはデータ提供者側での計算環境の提供の在り方等も絡んできますので、これらについても利用者ニーズあるいは提供者のありようを見つつ、今後検討していく必要があると考えるところでございます。

最後の●でございますが、利活用審査、いわゆるData Access Committee (DAC) でございますけれども、こちらにつきましては、データの特性や利用方法や形態、また、審査プロセスの迅速化、申請者へのプロセスの可視化等、利用者にとっての利便性を考慮し、在り方の検討を進め、一般利用の開始までには対応していく必要があると考えております。利活用審査につきましては、現状では一般的にはデータ提供者側の施設で行われることが多いと認識しておるところでございますが、多施設のデータ利用審査をどのように円滑に進めていくのであるとか、或いは審査機能を持たないデータ提供者についてはどのように対応していくのかといったことも今後検討の必要があると考えております。これらは府省の皆様方にも御指導を仰ぎつつ、検討を進めて参りたいと考えております。

以上のような構想を踏まえまして、この5ページを見ていただければと思います。こちらは2ページの図と全く同じものがございますが、上の枠を赤で囲っております。こちらは現在令和4年度予算によりまして開発を実施しているところでございます。この中で一部「データ持込み」と書かれた右側の下のほうでございますけれども、諸々の機能につきましても、細かくなりますが、今後令和5年度以降の開発としております。基本的な機能につきましては、今年度の調達で開発を行っていくところでございます。

最後に、6ページを御覧いただければと思います。これまで申し上げましたのはシステ

ム面でのお話でございますが、その他の事柄について御説明いたします。まず、左の四角でございますが、データ利活用に関する検討でございます。こちらは検討会を設置し、以下のようなデータ利活用に関する検討を進めております。すなわち、現在AMED研究を対象とした説明文章用モデル文案について検討しておりますけれども、こちらにつきましては、データの第三者提供という観点から先行して作成された厚生労働省における全ゲノム解析等実行計画の説明文書用モデル文案を参照しつつ我々も作成しており、下記のとおり検討会を重ねており、今週も第3回の会合を開いたところでございます。

また、右側でございますが、データ及び分析等の連携に関する検討、すなわちシステム面やサービス面の検討につきましては、先ほど申しましたとおり、現在連携基盤の構築についての調達を実施中でございます。加えまして、関係の諸機関と今後システムの開発や具体的な検討についての会合を行い、その会合も今後重ねていながら検討して参りたいと思います。また、有識者を交えた検討会につきましても、12月に第1回を開催したところでございます。このように、ソフト面におきましても検討会で御指導を仰ぎつつ進めて参りたいと考えております。

説明は以上になります。

○長野次長 ありがとうございます。

続きまして、先般11月の健康・医療のデータ利活用協議会で決められておりますデータ連携の進め方について、事務局より説明いたします。

○神田参事官 事務局の神田と申します。

資料3-2の説明をさせていただきます。

健康・医療研究開発データ統合利活用プラットフォーム、いわゆるCANNDsの進捗については、今ほどAMEDから説明をいただいたところですが、既存のバイオバンク事業、全ゲノム解析等実行計画、その他の事業とどのようにデータ連携を進めていくのかという全体像について、先月の健康・医療データ利活用基盤協議会において取りまとめた資料になります。これの御紹介をさせていただきます。

最初のパラグラフですけれども、CANNDsはどんなものかということですが、複数のデータベース等を連携し、横断検索機能を有するとともに、産業界も含めた研究開発に大規模なデータを扱える場を広く提供する、こういったものであることを述べております。今年度から予算事業化したCANNDsですけれども、先ほどのAMEDの説明のとおり、まずは先行的にいわゆる三大バイオバンク、全ゲノム解析等実行計画などにおいて収集されるゲノムデータをつなぐための開発を進めることとしております。また、ゲノムデータのみではなく、臨床情報も含む様々なデータとも連携することで活用の幅が広がってくるわけですので、これも引き続き進めることとしております。

CANNDsの基盤機能の開発につきましては、データ連携を前提として、国民や関係者から信頼感を得る、そのために安全・安心の確保に留意することが重要だという指摘をいただいております。そのことを記載しております。また、開発は段階的に行っていくわけで

すけれども、当初から柔軟性が高い相互運用性を持たせた開発を行います。基盤機能に含まれない特定用途のソフトウェア、これはデータを利活用する事業によっていろいろな利用ニーズがあるわけですが、そうした特定用途のソフトウェア開発については、それらの事業に委ねるということになります。

次のパラグラフですけれども、AMEDはいわゆる三大バイオバンクや全ゲノム解析等実行計画などの関係機関の参画を得ながら共通のシステム開発を進めます。特に、AMEDが全ゲノム解析等実行計画におけるがん・難病のデータを対象とした横断検索機能等の構築を行う際には、先行してCANNDsの基盤機能の開発を進めているわけですので、ユーザーインターフェースの一体化や共通の機能の提供を実現させるために、先行して開発した成果を活用するということになります。

令和5年度・6年度に開発を必要とする要素についても、AMEDは、現時点からデータ連携先の関係機関と議論を進めて参ります。今後連携が進むゲノム以外のデータにつきましては、社会的要請の高いものから、データの質を担保しつつ順次連携を実施していくこととなります。データ連携の拡充に伴う追加機能の開発については、当初開発をする基盤機能ができる限り活用することでコストを抑えるとともに迅速にシステムを改修していくこととしております。

これらのAMEDの取組には、内閣府、文科省、厚労省、経産省及び関係機関は最大限協力をするという事で取りまとめをしております。

2枚目のものは文章をポンチ絵で表現したものになります。

説明は以上です。

○長野次長 ありがとうございます。

それでは、ただいまのAMEDからの説明、事務局の説明につきまして、何か御質問、御意見などがございましたら、お時間としては5分ちょっとでございます。ございますでしょうか。

上野構成員、よろしく申し上げます。

○上野構成員 ありがとうございます。製薬協の上野でございます。

最初の3-1の資料の中で、被験者様からいただく同意書の管理について、細かくて具体的なお話になるかもしれませんが、要望をお願いしたいと思います。5ページにあります「データを扱う場の提供」の中で同意管理という項目がありまして、その同意書につきましては、最後のほうで述べられたモデルの文案をつくられているということで、いろいろな機関が違う中でこういった同意文案を一つの形にまとめるという点については、非常に我々としてもありがたい取組だと思います。ただ、その上で、それぞれの同意書をそれぞれの機関に広がっているものを私ども利用者がそれぞれアクセスして集めて確認するということだと非常に煩雑な作業になりますので、この同意管理というシステムがCANNDsの中にあるのであれば、AMEDの中でそういった同意文書を一元管理して利用者側が一元的にアクセスして使いやすい、そういった環境を整えていただければと思いますので、

ぜひ御検討いただければと思います。

以上です。

○長野次長 ありがとうございます。

AMEDからレスポンスはございますか。

○松岡課長 松岡から回答いたします。貴重な御意見をありがとうございます。

AMEDが支援する研究課題を全て把握しているわけではございませんが、およそ3,000課題、仮に協力機関が5機関存在するとしますと、1万5000種類ぐらいの説明文書が日本に津々浦々存在し得ることになります。ですから、我々は今、説明文書のモデル文案というものを作成しており、ある程度ここは確実に書いてくださいというところと、変更可能な箇所も当然存在します。ただ、変更可能な箇所が当然発生するはずなのですが、それを全てチェックすることは容易ではございません。したがって、モデル文案、今、作成をしておりますので、これを一切変更せずに使用したのか、それとも変更した箇所があるのかというような機械的な確認は可能かもしれませんが、いただいた御意見等も踏まえまして、運用につきましてはこちらの事業の運営委員会で検討させていただければと思います。

○上野構成員 よろしく願いいたします。

○長野次長 続きまして、菅野構成員、よろしく願いします。

○菅野構成員 最初の前半で非常にCANNDsに対する期待感が高かったと思うのですが、今のタイムコースを見るとそういう形になるのは5年後ぐらいかというタイムコースの感じで、これはかなりのんびりした感じかという気がしております。でも、一朝にして物事はできないので、順次やっていかざるを得ないと思いますが、その中でも具体的にAMEDにコンピューターを置いて窓口となって例えば来年度コントロールのシークエンスについてはサービスを開始するというような、そういう御計画なのでしょうか。それとも、今、スパコン連携という形で各省が持っていると思いますが、そこへ行ってもらおうというような実際の運用になるのでしょうか。

○松岡課長 御質問ありがとうございます。松岡から回答できるところを回答させていただきます。

現在こういった矢印に示しておりますように、決して素早くシステムができるということでは御座いませんで、利用開始を希望されている皆様方には誠に申し訳なく思っておりますけれども、确实なところから一步一步進めて参りたいと考えており、特に今年度の末から来年度はパイロット運用から限定提供、一般提供という、これはコントロール群、三大バイオバンクゲノムのデータの利活用を進めていくということでございます。こちらにつきまして、どのような体制で実際に運営していくかという運営面に関しましては様々な形態が想定されており、まだこちらはこれといった形で確定しておらず、そこも検討していく必要があるだろうと考えております。これも利用者の方々がどのようなニーズを持っていらっしゃるかということで、我々も製薬関連、バイオ、再生医療、また、食品や健康関連、ITやデータ解析関連の企業、団体から約50者の候補をリストアップしまして、19



者のヒアリングを終えております。内容としましては、想定されるニーズやユースケース、また、利用環境、利用コスト、利用申請に関する時間等、様々な意見を聞いておりますので、それらを今、まとめております。そういった御意見も踏まえまして、有り様については検討させていただければと思っております。

以上でございます。

○菅野構成員 ありがとうございます。

○長野次長 それでは、次に高木構成員、挙手されていますか。よろしく申し上げます。

○高木構成員 高木でございます。

CANNDsの横断検索といいたいでしょうか、機能についての質問が1件とコメントが1件ございます。まず質問でございますが、CANNDsで横断検索をされるわけですが、IDやメタデータが各バンクなりデータそれぞれで違ってきますと、実際には探せるものも探せないことになってきますので、その辺りの検索機能のためのメタデータの統一といいたいでしょうか、IDの統一といいたいでしょうか、そういうことをどうお考えになっているのか。それから、それぞれのデータのデータ提供者でフォーマットが違ってきますと、検索できた後もそれを組み合わせて使うことが非常に困難になるのではないかとも思うのですが、その辺りについてどうお考えなのか、これが1点目の質問でございます。まずはここで止めたいと思います。

○松岡課長 御質問ありがとうございます。

御指摘の点、大変重要でございます。こちらはデータ協議会の方でも類似した御質問をいただいております。恐らく横断検索等を行うにおいてバイオバンク、横断検索とどのような違いがあるのか、類似性があるのかということとも関連するかと思います。こちら我々は内閣府健康・医療戦略推進事務局、また、文部科学省とも御相談しております。なるべく同じ機能の再開発は極力しないということで、9月以降にAMED内の関係者が集まりまして議論しております。横断検索の開発者とも10月以降から意見交換を行っており、先ほど御指摘のございましたIDの連携等についても議論しております。具体的には、我々のプラットフォームで実施したいことがより明確になった時点で、さらにそれについても明確にしやすいだろうということから、具体を現在検討しております。ですから、現時点でこの場で明確な御回答はできませんけれども、そういった問題意識については我々も有しており、検討を進めているというところで御理解をいただければと思います。

○高木構成員 分かりました。ありがとうございます。

今、ちょうど出て参りましたけれども、2点目のコメントと申し上げたのは、仰ったバイオバンク横断検索や或いは他に関連するデータベースや検索機能がございまして、ぜひその辺りと連携を取った形で進めていただければと思います。それがコメントでございます。

以上です。

○松岡課長 ありがとうございます。

○長野次長 ありがとうございます。

時間も来てございますが、笠貫参与、岡田先生、これで最後にさせていただきます。

○笠貫参与 ありがとうございます。

実行計画2022、CANNDsが進んでいて、心強いと思います。2つコメントさせていただきます。

第一はデータサイエンスです。データサイエンティストを人材育成として入れていただきたいと思います。ゲノムオミックス情報に臨床表現型情報や環境情報、さらに時系列のオミックスデータなどが入りビッグデータになります。そうしますと、数理、情報学、疫学、統計学など、ビッグデータの理論や知見を統合する学問としてデータサイエンスが非常に重要です。第二は、PPIの話が出ましたが、日本では2005年に既に科学技術コミュニケーション元年を宣言しています。その時、東大で科学技術インタープリター、北大で科学技術コミュニケーター、そして、早稲田大学で科学技術ジャーナリストが養成されています。そうした科学技術コミュニケーション政策の展開の流れでこの実行計画を考えていただきたいということです。

もう一つ、個別化医療の言葉ですが、プレジジョンメディシンと区別して使われているのかを考えていただけたらと思います。

以上です。

○長野次長 ありがとうございます。

では、岡田先生からコメントがございませうか。

○岡田参考人 手短かに。CANNDsのお話ですけれども、現場的な話をいたしますと、CANNDsがあるのでデータを公開しませんというロジックを最近よく聞くようになってきております。つまり、CANNDsというのはデータの解析のプラットフォームであって、データを公開するNBDCなどを置き換えるものではないという位置づけだと思うのです。なのですけれども、CANNDsにデータを出すのでNBDCには出しませんというようなことを言う人が結構増えてきておまして、それはCANNDsがまだ発展途上であることを当然含めて言っているわけですね。ですから、こういう枠組みと、それはそれとして元データをデポジットしていくことを並行して行っていくということを周知していただければ幸いです。

○長野次長 貴重な情報をありがとうございます。 それでは、本日の議題、以上となりますけれども、事務局から今後の予定などについて連絡をお願いします。

○笠松参事官 事務局でございます。

次回の協議会につきましては、日程等、改めて事務局から御連絡を差し上げたいと思います。

また、過去の協議会で御指摘いただいておりますけれども、ゲノム医療推進のための研究開発における全般的な課題、これをトータルで議論する場を持たせていただきたいと思っておりますので、また御相談させていただきます。

以上でございます。

○長野次長 それでは、これもちまして第10回ゲノム医療協議会を終了いたします。本日は御多忙のところ、皆様どうもありがとうございました。